

Ebaluatoia 2014: ingelese-euskara itzultzaile automatikoen konparazioa

Nora Aranberri, Arantza Diaz de Ilarraza, Gorka Labaka, Kepa Sarasola
Ixa taldea (EHU) <http://ixa.si.ehu.es/tacardi>

Azken hamarkadan Itzulpen Automatikorako (IA) ebaluazio-neurririk erabiliena bilakatu den Bleu (Papineni *et al.*, 2002) metrikaren inguruan zalantza ugari sortu izan dira (Callison-Burch *et al.*, 2006). Itzultzaile estatistiko baten garapenean lagungarri izan daiteke Bleu metrika, hobekuntzen traza jarraitu ahal izateko, baina ez du IAKo sistemen itzulpenaren kalitate absolutua neurtzen, ezta erabiltzaile batentzat itzulpenak zenbateraino diren baliagarriak erakusten. Gainera estrategia desberdinekin sortutako sistemak konparatzeko ere ez du balio (Mayor *et al.*, 2009). Bestalde eskuzko giza ebaluazioak oso garestiak izaten dira eta antolatzeke zailak. Hizkuntza teknologiaren barruan, baina, ebaluazio-kanpaina asko egin izan dira. Adibidez, Workshop on Statistical Machine Translation barruan 2006tik urtero egin da horrelako bat, baina partehartzaileak gehienetan ikerlariak eurak izan dira. Kanpaina horietan Amazon-en Mechanical Turk bezalako plataformak erabili ohi dira.

2013an ENEUS proiektuaren barruan IXA taldeak ingelesetik euskarara automatikoki itzultzen zuten lau itzultzaile automatiko garatu zituen. Lehenengo prototipo haietako bakoitza teknologia desberdinekin eraiki ziren: 1) MATXIN ENEUS: erregelatan oinarritutako sistema; 2) SMTb: oinarritzko sistema estatistikoa, 3) SMTs: estatistikoa baina hitzen segmentazioa erabilita, eta 4) Matxin eta sistema estatistiko hoberena batzen zituen sistema *hibridoa*. Eta proiektuaren bukaeran ebaluatu egin behar ziren.

Batzuek zioten euskara landuko zuen ebaluatoi masiborik ez zela inoiz ondo aterako, euskararen mundua txikiegia zela, baina IXA taldean animatu ginen 2014.eko otsailean horrelako ebaluatoi bat antolatzen eta gure harridurarako erantzun ikaragarria lortu genuen. Zer espero genuen? Zelako erantzuna jaso genuen? Zein izan zen ebaluatoiaren emaitza ingelesetik euskarara zebiltzan orduko bost itzultzaileen artean? Erraza litzateke esperimendu arrakastatsu hau inoiz errepikatzea? Galdera horiei erantzun nahi diegu.

Zer espero genuen?

10.000 ebaluazio prestatu genituen. Hori lortzea zen gure ametsa. Gutxienez 250 esaldi ebaluatzea beharrezkoa dela esan ohi da emaitza esanguratsuak lortu nahi badira. Honek esan nahi du, gure kasuan, non 5 sistema (Google Translate ere bai) konparatuko ziren eta kontrol esaldiak (ausaz emandako erantzunak detektatu eta ekiditekoak) jarriko genituen, 10.715 ebaluazio inguru bildu behar genituela (7.500 esaldi guztiak sistema pare guztietarako 3 aldiz ebaluatuta izateko eta gainontzeko 3.215 kontrolerarako).

Zelako erantzuna jaso genuen?

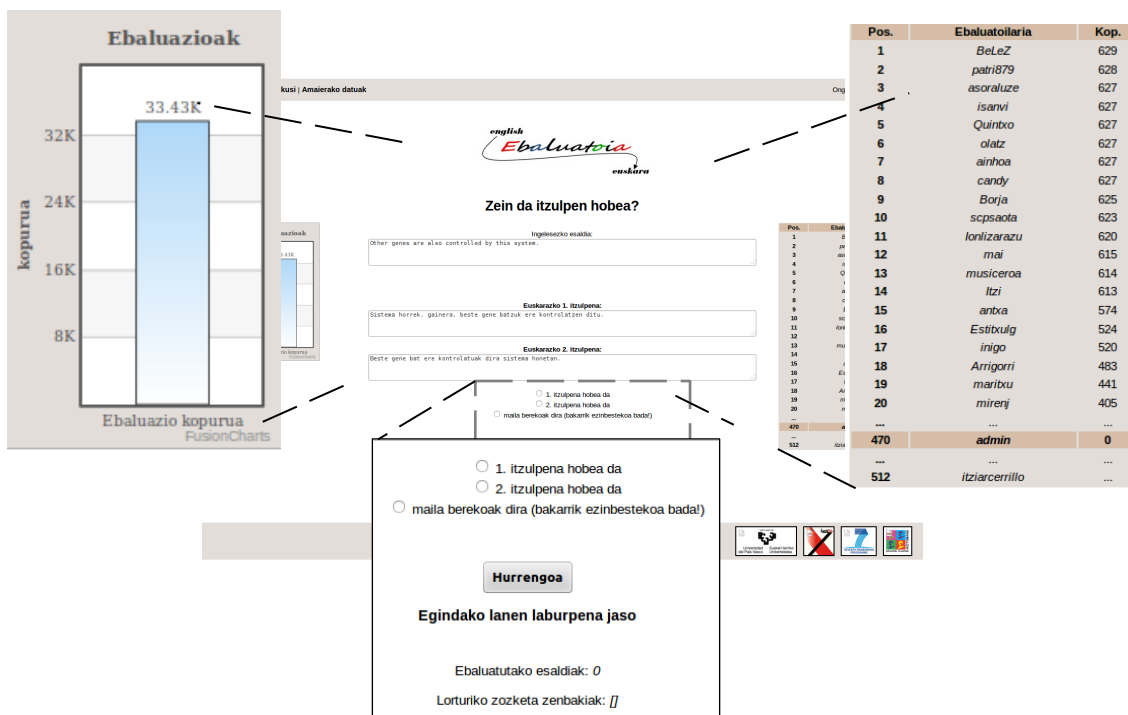
Guztira 33.430 ebaluazio jaso genituen, 512 parte-hartzailek eginda. Bigarren egunerako gaituta zegoen 10.000 ebaluazioko helburua.

Zein izan zen ebaluatoiaren emaitza ingelesetik euskarara zebiltzan orduko bost itzultzaileen artean?

Language Resources and Evaluation aldizkarian onartuta eta argitaratzeaz dugun azterketetan (Aranberri *et al.*, 2015) sakonki azaltzen denez, ebaluatoian jasotako datuetatik sistema itzultzaileen sailkapena, hoberenetik txarrenera, hau izan zen: *SMTs* \approx *Google* > *SMTb* > *Hybrid* > *Matxin ENEUS*. Soilik neurri estatistiko automatikoak erabilita, aldiz, beste sailkapen hau aterazten zen: *Google* > *SMTs* > *SMTb* > *Hybrid* > *Matxin ENEUS*. Beraz, neurri estatistikoek esaten zuten Google-ren itzultzailea zela hoberena dudarik gabe, baina pertsonak esan zuten hitzen segmentazioa erabiltzen zuten SMTs sistema estatistikoa eta Google berdintsu zebiltzala. Pena Eusko Jaurlaritzako *Itzultzailea* ez genuela ebaluatu, esperimendua diseinatu genuenean ez baitzegoen martxan.

Erraza litzateke esperimendu arrakastatsu hau inoiz errepikatzea?

Ez, ez litzateke erraza. Ebaluazio-esperimendu honetan inbertitu genituen baliabideak handiak izan ziren. Egia da neurri handi batean berrerabili ahal izango genukeela 2014an Elhuyarrekoeekin sortu genuen aplikazio informatikoa (Mechanical Turk baliatzea ez zen posible izan gure esperimenduan). Baina lan ikaragarriak izan ziren (1) ebaluaziorako hizkuntza-baliabideak lortzea, (2) prozesuan sortutako arazo teknikoak denbora errealean konpontzea, eta (3) sare sozialen bidez prozesua dinamizatzea. Hilero ezin dugu errepikatu horrelako ebaluazio bat, ezta urtero ere.



Erreferentziak

- Alegria I., Díaz de Ilarraza A., Labaka G., Lersundi M., Sarasola K. 2008. [Itzulpen automatikoa: aukerak, arazoak eta erronkak](#). *Bat Soziolinguistika*. 66. zenb. 107-122 orr. ISSN:1130-8435
- Nora Aranberri, Gorka Labaka, Arantza Díaz de Ilarraza, Kepa Sarasola. 2015. *Ebaluatoia: crowd evaluation for English-Basque machine translation*. *Language Resources and Evaluation Journal*. (Argitaratzeaz)
- Nora Aranberri. 2014. **Itzulpen automatikorako Matxin sistemaren portabilitate-azterketa ingeles-euskara prototipoa lagun EUSKAL HIZKUNTZALARITZAREN EGUNGO ZENBAIT IKERLERRO. HIZKUNTZALARI EUSKALDUNEN I. TOPAKETA**. Itziar Aduriz eta Ruben Urizar (ed.) *Udako Euskal Unibertsitatea Bilbo, 2014 ISBN: 978-84-8438-524-0 Lege-gordailua: BI-1399-2014*
- Callison-Burch C., Osborne M., and Koehn P. Re-evaluating the role of BLEU in *Machine Translation research*. In *Proceedings of EACL-2006*, 2006.
- Hizkuntza teknologia bloga, 2014 [Ebaluatoia, ingelesetik euskarara itzultzen duten lehen prototipoak ebaluatzeo](#) <http://www.unibertsitatea.net/blogak/ixa/2014/02/12/ebaluatoia-ingelesetik-euskarara-itzultzen-duten-lehen-prototipoak-ebaluatzeo/>
- Hizkuntza teknologia bloga, 2014 [Bukatu da Ebaluatoia. Parte hartzea itzela izan da](#). <http://www.unibertsitatea.net/blogak/ixa/2014/03/05/bukatu-da-ebaluatoia-parte-hartzea-itzela-izan-da/>
- Aingeru Mayor, Iñaki Alegria, Arantza Díaz de Ilarraza, Gorka Labaka, Mikel Lersundi, Kepa Sarasola 2009 **Matxin, euskararako lehenengo itzultzaile automatikoa SENEZ**. *Itzulpen Aldizkaria*, 37 zk. (197--220 orr) ISSN:1132-2152. <http://www.eizie.org/Argitalpenak/Senez/20091026/matxin>
- Papineni K., Roukos S., Ward T., and Zhu W. BLEU: a method for automatic evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311-318, 2002.